Amendments to the claims:

This listing of claims will replace all prior versions, and listings, of claims in the application:

Listing of Claims:

1. (Currently amended) A method ~~of crawling documents~~ comprising:

extracting a set of uniform resource locators (URLs) from at least one document;

analyzing the ~~extracted~~ set of URLs extracted from the at least one document to determine those in the set of URLs that contain session identifiers by locating the session identifiers in the set of URLs extracted as sub-strings that occur in multiple URLs of a web site;

generating a clean set of URLs from the ~~extracted~~ set of URLs extracted from the at least one document by removing ~~using~~ the session identifiers; and

determining when at least one ~~second~~ particular URL has already been crawled based, at least in part, on a comparison of the ~~second~~ particular URL to the clean set of URLs.

2. (Cancelled)

3. (Original) The method of claim 1, wherein the at least one document is a web document downloaded from a web site.

4. (Currently amended) The method of claim 1, wherein the comparison of the ~~second~~ particular URL to the clean set of URLs is based on a comparison of a

fingerprint value calculated for each of the URLs in the clean set of URLs.

5. (Original) The method of claim 1, wherein the session identifiers are determined as including sub-strings from the set of URLs that do not reference content.

6. (Cancelled)

7. (Currently amended) The method of claim [[6]] 1, wherein the analyzing the ~~extracted~~ set of URLs extracted from the at least one document further includes:

locating the session identifiers in the extracted set of URLs as sub-strings that contain characters consistent with a session identifier.

8. (Currently amended) The method of claim 1, further comprising:

downloading content from the ~~second~~ particular URL when the ~~second~~ particular URL is determined to not already have been crawled.

9. (Currently amended) The method of claim 1, further comprising:

storing information based on the clean set of URLs for use in later determining whether additional URLs have already been extracted; and

storing the ~~extracted~~ set of URLs extracted from the at least one document, including embedded session identifiers, for use in later accessing the ~~extracted~~ set of URLs extracted from the at least one document.

10. (Currently amended) A method comprising:

receiving a set of uniform resource locators (URLs);

analyzing the set of URLs for sub-strings that are structured in a manner consistent with session identifiers; and

further analyzing the set of URLs to identify those of the sub-strings as corresponding to session identifiers based on multiple occurrences of a sub-string in the set of URLs.

11. (Original) The method of claim 10, wherein the set of URLs are extracted from a web document associated with a web host.

12. (Original) The method of claim 10, wherein the set of URLs are extracted from multiple web documents associated with a single web host.

13. (Original) The method of claim 10, further comprising:

removing identified session identifiers from the set of URLs; and

storing the set of URLs with the removed session identifiers as a clean set of URLs.

14. (Currently amended) The method of claim 13, further comprising:

adding a generated session identifier to URLs in the clean set of URLs ~~when the URLs are to be used to access a web document~~.

15. (Currently amended) A device comprising:

at least one fetch bot configured to download content on a network from locations specified by uniform resource locators (URLs);

a content manager configured to

extract URLs from the downloaded content, and

identify session identifiers from the ~~extracted~~ URLs extracted from the downloaded content based, at least in part, on multiple occurrences of the session identifiers from a single web site; and

a URL manager configured to store clean versions of the ~~extracted~~ URLs extracted from the downloaded content in which the session identifiers are removed from the ~~extracted~~ URLs extracted from the downloaded content.

16. (Currently amended) The device of claim 15, wherein the content manager is further configured to identify the session identifiers based on locating sub-strings, within the URLs extracted from the downloaded content, that contain characters consistent with session identifiers.

17. (Original) The device of claim 15, further comprising:

a database configured to store the downloaded content.

18. (Currently amended) The device of claim 15, wherein the URL manager is further configured to determine when additional URLs have previously been stored by

comparing clean versions of the additional URLs to the stored clean versions of the ~~extracted~~ URLs extracted from the downloaded content.

19. (Currently amended) The device of claim 15, wherein the session identifiers include characters from the ~~extracted~~ URLs extracted from the downloaded content that do not reference content.

20. (Currently amended) A device comprising:

means for receiving a set of uniform resource locators (URLs);

means for analyzing the set of URLs for sub-strings that are structured in a manner consistent with session identifiers; and

means for further analyzing the set of URLs to identify those of the sub-strings as corresponding to session identifiers based on multiple occurrences of a sub-string in the set of URLs.

21. (Original) The device of claim 20, wherein the set of URLs are extracted from a web document associated with a web host.

22. (Original) The device of claim 20, wherein the set of URLs are extracted from multiple web documents associated with a single web host.

23. (Original) The device of claim 20, further comprising:

means for removing the identified session identifiers from the set of URLs; and

means for storing the set of URLs with the removed session identifiers as a clean set of URLs.

24. (Currently amended) The device of claim 23, further comprising:

means for adding a generated session identifier to URLs in the clean set of URLs ~~when the URLs are to be used to access a web document~~.

25. (Currently amended) A computer-readable medium including programming instructions that when executed by at least one processor causes the at least one processor to perform a method including:

receiving a set of uniform resource locators (URLs);

analyzing the set of URLs for sub-strings that are structured in a manner consistent with session identifiers; and

further analyzing the set of URLs to identify those of the sub-strings as corresponding to session identifiers based on multiple occurrences of a sub-string in the set of URLs.

26. (Original) The computer-readable medium of claim 25, wherein the set of URLs are extracted from a web document associated with a web host.

27. (Original) The computer-readable medium of claim 25, wherein the set of URLs are extracted from multiple web documents associated with a single web host.

28. (Original) The computer-readable medium of claim 25, wherein the programming instructions further include programming instructions that cause the at least one processor to:

remove the session identifiers from the set of URLs; and

store the set of URLs with the removed session identifiers as a clean set of URLs.

29. (Original) The computer-readable medium of claim 28, wherein the programming instructions further include programming instructions that cause the at least one processor to:

add a generated session identifier to URLs in the clean set of URLs when the URLs are to be used to access a web document.